NOVEL DNA SEQUENCES

TECHNICAL FIELD

This invention relates to novel DNA constructs encoding proteolytic enzymes, as well as recombinant expressions vectors and host cells comprising these DNA constructs, and methods of producing a proteolytic enzyme.

BACKGROUND ART

WO 88/03947 describes a novel alkaline protease prepared by cultivating a strain of *Nocardiopsis sp.*, and its use in detergent compositions.

WO 93/13193 describes the use of proteases derived from members of the genus *Nocardiopsis* in detergent additives or compositions, or wash liquors, comprising specific bleaching systems.

Although proteolytic enzymes obtained by cultivating a strain of *Nocardiopsis sp.* have been described, their amino acid sequences or DNA sequences encoding these enzymes have never been disclosed.

SUMMARY OF THE INVENTION

According to the present invention, the inventors have now succeeded in isolating and characterizing a DNA sequence encoding a proteolytic enzyme, thereby making it possible to prepare a mono-component enzyme preparation.

Therefore, in its first aspect, the invention provides a DNA construct comprising a DNA sequence encoding a proteolytic enzyme, which DNA sequence,

(a) comprises the DNA sequence presented as SEQ ID NO: 1; or

(b) comprises a sequence analogue to the DNA sequence presented as SEQ ID NO: 1, which analog DNA sequence either

(i) is homologous to the DNA sequence presented as SEQ ID NO: 1; or

(ii) hybridizes with the same oligonucleotide probe as the DNA sequence presented as SEQ ID NO: 1; or

(iii) encodes a polypeptide which is at least 70% homologous with the polypeptide encoded by the DNA sequence presented as SEQ ID NO: 1; or

(iv) encodes a polypeptide which is immunologically reactive with an antibody raised against a purified protease derived from the strain *Nocardiopsis sp.* 10R NRRL 18262, or encoded by the DNA sequence presented as SEQ ID NO: 1.

In further aspects the invention provides a recombinant expression vector comprising the DNA construct of the invention, as well as a cell comprising the DNA construct of the invention or the recombinant expression vector of the invention.

Finally the invention provides a method of producing a proteolytic enzyme, the method comprising culturing the cell of the invention under conditions permitting the production of the enzyme, and recovering the enzyme from the culture, as well as a proteolytic enzyme, which is encoded by a DNA construct of the invention, is produced by the method of the invention, and/or is immunologically reactive with an antibody raised against a purified protease derived from the strain *Nocardiopsis sp.* 10R NRRL 18262, or encoded by the DNA sequence presented as SEQ ID NO: 1.

DETAILED DISCLOSURE OF THE INVENTION

DNA Constructs

The present invention provides a DNA construct comprising a DNA sequence encoding a proteolytic enzyme, which DNA sequence,

(a) comprises the DNA sequence presented as SEQ ID NO: 1; or

(b) comprises a sequence analogue to the DNA sequence presented as SEQ ID NO: 1, which analog DNA sequence either

(i) is homologous to the DNA sequence presented as SEQ ID NO: 1; or

(ii) hybridizes with the same oligonucleotide probe as the DNA sequence presented as SEQ ID NO: 1; or

(iii) encodes a polypeptide which is at least 70% homologous with the polypeptide encoded by the DNA sequence presented as SEQ ID NO: 1; or

(iv) encodes a polypeptide which is immunologically reactive with an antibody raised against a purified protease derived from the strain *Nocardiopsis sp.* 10R NRRL 18262, or encoded by the DNA sequence presented as SEQ ID NO: 1.

As defined herein the term “DNA construct” is intended to indicate any nucleic acid molecule of cDNA, genomic DNA, synthetic DNA or RNA origin. The term "construct" is intended to indicate a nucleic acid segment which may be single- or

double-stranded, and which may be based on a complete or partial naturally occurring nucleotide sequence encoding the proteolytic enzyme of interest. The construct may optionally contain other nucleic acid segments.

The DNA construct of the invention encoding the protease may suitably be of genomic or cDNA origin, for instance obtained by preparing a genomic or cDNA library and screening for DNA sequences coding for all or part of the protease by hybridization using synthetic oligonucleotide probes in accordance with standard techniques (cf. e.g. *Sambrook et al.*, Molecular Cloning. A Laboratory Manual, Cold Spring Harbor, NY, 1989).

The nucleic acid construct of the invention encoding the protease may also be prepared synthetically by established standard methods, e.g. the phosphoamidite method described by *Beaucage and Caruthers*, Tetrahedron Letters 1981 **22** 1859-1869, or the method described by *Matthes et al.*, EMBO Journal 1984 **3** 801-805. According to the phosphoamidite method, oligonucleotides are synthesized, e.g. in an automatic DNA synthesizer, purified, annealed, ligated and cloned in suitable vectors.

Furthermore, the nucleic acid construct may be of mixed synthetic and genomic, mixed synthetic and cDNA or mixed genomic and cDNA origin prepared by ligating fragments of synthetic, genomic or cDNA origin (as appropriate), the fragments corresponding to various parts of the entire nucleic acid construct, in accordance with standard techniques.

The nucleic acid construct may also be prepared by polymerase chain reaction using specific primers, for instance as described in US 4,683,202 or by *Saiki et al.*, Science 1988 **239** 487-491.

In a currently preferred embodiment, the nucleic acid construct of the invention comprises the DNA sequence shown in SEQ ID NO: 1, or any subsequence thereof, but which differ from the DNA sequence shown in SEQ ID NO: 1 by virtue of the degeneracy of the genetic code. The invention further encompasses nucleic acid sequences which hybridize to a nucleic acid molecule (either genomic, synthetic or cDNA or RNA) encoding the amino acid sequence shown in SEQ ID NO: 1, or any subsequence thereof, under the conditions described below.

Analogous DNA Sequences

As defined herein, a DNA sequence analogue to the DNA sequence presented as SEQ ID NO: 1 is intended to indicate any DNA sequence encoding a proteolytic enzyme, which enzyme has one or more of the properties cited under (i)-(iv), above.

The analogous DNA sequence may be isolated from another or related (e.g. the same) organism producing the protease on the basis of the DNA sequence presented as SEQ ID NO: 1, or any subsequence thereof, e.g. using the procedures described herein, and thus, e.g. be an allelic or species variant of the DNA sequence comprising the DNA sequence presented herein.

Alternatively, the analogous sequence may be constructed on the basis of the DNA sequence presented as SEQ ID NO: 1, or any subsequence thereof, e.g. by introduction of nucleotide substitutions which do not give rise to another amino acid sequence of the protease encoded by the DNA sequence, but which corresponds to the codon usage of the host organism intended for production of the enzyme, or by introduction of nucleotide substitutions which may give rise to a different amino acid sequence.

When carrying out nucleotide substitutions, amino acid changes are preferably of a minor nature, that is conservative amino acid substitutions that do not significantly affect the folding or activity of the protein, small deletions, typically of one to about 30 amino acids; small amino- or carboxyl-terminal extensions, such as an amino-terminal methionine residue, a small linker peptide of up to about 20-25 residues, or a small extension that facilitates purification, such as a poly-histidine tract, an antigenic epitope or a binding domain. Examples of conservative substitutions are within the group of basic amino acids (such as arginine, lysine, histidine), acidic amino acids (such as glutamic acid and aspartic acid), polar amino acids (such as glutamine and asparagine), hydrophobic amino acids (such as leucine, isoleucine, valine), aromatic amino acids (such as phenylalanine, tryptophan, tyrosine) and small amino acids (such as glycine, alanine, serine, threonine, methionine). For a general description of nucleotide substitution, see e.g. *Ford et al.*, Protein Expression and Purification, 2 1991 95-107.

It will be apparent to persons skilled in the art that such substitutions can be made outside the regions critical to the function of the molecule and still result in an active polypeptide. Amino acids essential to the activity of the polypeptide encoded by the DNA construct of the invention, and therefore preferably not subject to substitution, may be identified according to procedures known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (cf. e.g. *Cunningham and Wells*; Science 1989 **244** 1081-1085). In the latter technique mutations are introduced at every residue in the molecule, and the resultant mutant molecules are tested for biological (i.e. proteolytic) activity to identify amino acid residues that are critical to the activity of the molecule. Sites of substrate-enzyme interaction can also be determined by analysis of crystal structure as determined by such techniques as nuclear magnetic

resonance analysis, crystallography or photoaffinity labeling (cf. e.g. *de Vos et al.*; Science 1002 **255** 306-312; *Smith et al.*; J. Mol. Biol. 1992 **224** 899-904; *Wlodaver et al.*; FEBS Lett. 1992 **309** 59-64).

It will be understood that the DNA sequence presented as SEQ ID NO: 1, or any subsequence thereof may be used as probes for isolating the entire DNA sequence encoding the proteolytic enzyme.

The homology referred to in (i) above is determined as the degree of identity between the two sequences indicating a derivation of the first sequence from the second. The homology may suitably be determined by means of computer programs known in the art such as GAP provided in the GCG program package (*Needleman S B & Wunsch C D*; J. Mol. Biol. 1970 **48** 443-453). Using GAP with the following settings for DNA sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the DNA sequence exhibits a degree of identity preferably of at least 70%, in particular at least 80%, or at least 85%, or at least 90%, or at least 95%, to the coding region of the DNA sequence shown in SEQ ID NO: 1.

The hybridization referred to in (ii) above is intended to indicate that the analogous DNA sequence hybridizes to the same probe as the DNA sequence encoding the protease under certain specified conditions which are described in detail in the Materials and Methods section hereinafter. The test for hybridization preferably is carried out under the conditions defined for low to medium stringency. In a more preferred embodiment, the test for hybridization preferably is carried out under the conditions defined for high stringency.

Normally, the analogous DNA sequence is highly homologous to the DNA sequence such as at least 70% homologous to the DNA sequence presented as SEQ ID NO: 1 encoding a protease of the invention, in particular at least 80%, or at least 85%, or at least 90%, or at least 95% homologous to said DNA sequence.

The degree of homology referred to in (iii) above is determined as the degree of identity between the two sequences indicating a derivation of the first sequence from the second. The homology may suitably be determined by means of computer programs known in the art. In a preferred embodiment the homology may be determined using the GAP program provided in the GCG program package (*Needleman S B & Wunsch C D*; J. Mol. Biol., 1970 **48** 443-453). Using GAP with the following settings for polypeptide sequence comparison: GAP creation penalty of 3.0 and GAP extension penalty of 0.1, the polypeptide encoded by an analogous DNA sequence exhibits a degree of identity preferably of at least 70%, in particular at least

80%, or at least 85%, or at least 95%, to the enzyme encoded by a DNA construct comprising the DNA sequence shown in SEQ ID NO: 1.

The term "derived from" in connection with property (iv) above is intended not only to indicate a protease produced by the strain *Nocardiopsis sp.* 10R NRRL 18262, but also a protease encoded by a DNA sequence isolated from this strain and produced in a host organism transformed with said DNA sequence. The immunological reactivity may be determined by the method described in the Materials and Methods section below.

The DNA sequence encoding an enzyme exhibiting proteolytic activity may be isolated by any general method involving

-cloning, in suitable vectors, a cDNA library, e.g. from the strain *Nocardiopsis sp.* 10R NRRL 18262,

-transforming suitable yeast host cells with said vectors,

-culturing the host cells under suitable conditions to express any enzyme of interest encoded by a clone in the cDNA library,

-screening for positive clones by determining any proteolytic activity of the enzyme produced by such clones, and

-isolating the enzyme encoding DNA from such clones.

A general method has been disclosed in WO 93/11249 or WO 94/14953, the contents of which are hereby incorporated by reference.

Microbial Sources

It is at present contemplated that a DNA sequence encoding an enzyme homologous to the enzyme encoded by the DNA sequence presented as SEQ ID NO: 1, i.e. an analogous DNA sequence, may be obtained from other microorganisms. For instance, the DNA sequence may be derived by screening a cDNA library of another microorganism, preferably a strain belonging to the order *Actinomycetes*, in particular a strain of *Nocardiopsis*.

Microorganisms belonging to the actinomycete *Nocardiopsis* are well known in the literature. Some examples of species and strains described are *Nocardiopsis dassonvillei*, Type Strain ATCC 23218; *Nocardiopsis dassonvillei* M58-1 (NRRL 18133), WO Pat. Publ. 88/03947; *Nocardiopsis dassonvillei* ZIMET 43647, DD Pat. Publ. 200,432; *Nocardiopsis dassonvillei* subsp. prasina,(Agric. Biol. Chem. 1990 **54**, 8, 2177-79); *Nocardiopsis sp.* OPC 120, JP Pat. Appl. 2,255,081; and *Nocardiopsis sp.* 10R (NRRL 18262), WO Pat. Publ. 88/03947.

Proteases derived from members of the actinomycete *Nocardiopsis* are disclosed in e.g. International Patent Application WO 88/03947 and GDR Patent No. DD 200,432.

Preferably, the proteases are derived from a protease producing strain of *Nocardiopsis dassonvillei*, preferably the strain ZIMET 43647, more preferred the strain *Nocardiopsis dassonvillei* M58-1 (NRRL 18133), or from a protease producing strain of the species defined by the strain 10R, more preferred the strain *Nocardiopsis sp.* 10R (NRRL 18262).

The strain *Nocardiopsis dassonvillei* ZIMET 43647 is described in the above mentioned DD Patent No. 200,432.

In a preferred embodiment, the DNA sequence encoding the protease is isolated by screening a cDNA library of the strain *Nocardiopsis sp.* 10R NRRL 18262. The strain *Nocardiopsis sp.* 10R NRRL 18262 has been deposited under the terms of the Budapest Treaty on 10 November 1987, at the Agricultural Research Service Culture Collection (NRRL), 1815 North University Street, Peoria, Illinois 61604, USA.

Being an International Depository Authority under the Budapest Treaty, NRRL affords permanence of the deposit in accordance with the rules and regulations of said treaty, *vide* in particular Rule 9. Access to the deposit will be available during the pendency of this patent application to one determined by the Commisioner of the United States Patent and Trademark Office to be entitled thereto under 37 C.F.R. Par. 1.14 and 35 U.S.C. Par. 122. Also, the above mentioned deposit fulfills the requirements of European patent applications relating to micro-organisms according to Rule 28 EPC.

DNA encoding the protease of the invention may, in accordance with well-known procedures, conveniently be isolated from DNA from a suitable source, such as any of the above mentioned organisms, by use of synthetic oligonucleotide probes prepared on the basis of a DNA sequence disclosed herein. For instance, a suitable oligonucleotide probe may be prepared on the basis of any of the nucleotide sequences presented as SEQ ID NO: 1, or any suitable subsequence thereof. A more detailed description of the screening method is given in Example 1 below.

Recombinant Expression Vectors

In another aspect, the invention provides a recombinant expression vector comprising the DNA construct of the invention.

The expression vector of the invention may be any expression vector that is conveniently subjected to recombinant DNA procedures, and the choice of

vector will often depend on the host cell into which it is to be introduced. Thus, the vector may be an autonomously replicating vector, i.e. a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g. a plasmid. Alternatively, the vector may be one which, when introduced into a host cell, is integrated into the host cell genome and replicated together with the chromosome(s) into which it has been integrated.

In the expression vector of the invention, the DNA sequence encoding the protease preferably is operably linked to additional segments required for transcription of the DNA. In general, the expression vector is derived from plasmid or viral DNA, or may contain elements of both. The term, "operably linked" indicates that the segments are arranged so that they function in concert for their intended purposes, e.g. transcription initiates in a promoter and proceeds through the DNA sequence coding the protease.

Thus, in the recombinant expression vector of the invention, the DNA sequence encoding the protease should be operably connected to a suitable promoter and terminator sequence. The promoter may be any DNA sequence which shows transcriptional activity in the host cell of choice and may be derived from genes encoding proteins either homologous or heterologous to the host cell. The procedures used to ligate the DNA sequences coding the protease, the promoter and the terminator, respectively, and to insert them into suitable vectors are well known to persons skilled in the art (cf. e.g. *Sambrook et al.*, Molecular Cloning. A Laboratory Manual, Cold Spring Harbor, NY, 1989).

The promoter may be any DNA sequence which shows transcriptional activity in the host cell of choice and may be derived from genes encoding proteins either homologous or heterologous to the host cell.

Examples of suitable promoters for directing the transcription of the DNA encoding the protease of the invention in bacterial host cells include the promoter of the *Bacillus stearothermophilus* maltogenic amylase gene, the *Bacillus licheniformis* alpha-amylase gene, the *Bacillus amyloliquefaciens* BAN amylase gene, the *Bacillus subtilis* alkaline protease gen, or the *Bacillus pumilus* xylanase or xylosidase gene, or by the phage Lambda P_R or P_L promoters or the *E. coli* lac, trp or tac promoters.

Examples of suitable promoters for use in yeast host cells include promoters from yeast glycolytic genes (*Hitzeman et al.*, J. Biol. Chem. 255 (1980), 12073 - 12080; *Alber and Kawasaki*, J. Mol. Appl. Gen. 1 (1982), 419 - 434) or alcohol dehydrogenase genes (*Young et al.*, in Genetic Engineering of Microorganisms for

Chemicals (*Hollaender et al*, eds.), Plenum Press, New York, 1982), or the TPI1 (US 4,599,311) or ADH2-4c (*Russell et al.*, Nature 304 (1983), 652 - 654) promoters.

Examples of suitable promoters for use in filamentous fungus host cells are, for instance, the ADH3 promoter (*McKnight et al.*, The EMBO J. 4 (1985), 2093 - 2099) or the tpiA promoter. Examples of other useful promoters are those derived from the gene encoding *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral α-amylase, *Aspergillus niger* acid stable α-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (gluA), *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase or *Aspergillus nidulans* acetamidase. Preferred are the TAKA-amylase and gluA promoters.

The expression vector of the invention may further comprise a DNA sequence enabling the vector to replicate in the host cell in question. The expression vector may also comprise a selectable marker, e.g. a gene the product of which complements a defect in the host cell, such as the gene coding for dihydrofolate reductase (DHFR) or the *Schizosaccharomyces pombe* TPI gene (described by *Russell P R*, Gene 1985 **40** 125-130), or one which confers resistance to a drug, e.g. ampicillin, kanamycin, tetracyclin, chloramphenicol, neomycin, hygromycin or methotrexate. For filamentous fungi, selectable markers include amdS, pyrG, argB, niaD and sC.

To direct the protease into the secretory pathway of the host cells, a secretory signal sequence (also known as a leader sequence, prepro sequence or pre sequence) may be provided in the expression vector. The secretory signal sequence is joined to the DNA sequence encoding the protease in the correct reading frame. Secretory signal sequences are commonly positioned 5' to the DNA sequence encoding the protease. The secretory signal sequence may be that normally associated with the protease or may be from a gene encoding another secreted protein.

In a preferred embodiment, the expression vector of the invention may comprise a secretory signal sequence substantially identical to the secretory signal encoding sequence of the *Bacillus licheniformis* α-amylase gene, e.g. as described in WO 86/05812.

Also, measures for amplification of the expression may be taken, e.g. by tandem amplification techniques, involving single or double crossing-over, or by multicopy techniques, e.g. as described in US 4,959,316 or WO 91/09129. Alternatively the expression vector may include a temperature sensitive origin of replication, e.g. as described in EP 283,075.

Procedures for ligating DNA sequences encoding the protease, the promoter and optionally the terminator and/or secretory signal sequence, respectively, and to insert them into suitable vectors containing the information necessary for replication, are well known to persons skilled in the art (cf. e.g. *Sambrook et al.*, Molecular Cloning. A Laboratory Manual, Cold Spring Harbor, NY, 1989).

Host Cells

In yet another aspect the invention provides a host cell comprising the DNA construct of the invention and/or the recombinant expression vector of the invention.

The DNA construct of the invention may be either homologous or heterologous to the host in question. If homologous to the host cell, i.e. produced by the host cell in nature, it will typically be operably connected to another promoter sequence or, if applicable, another secretory signal sequence and/or terminator sequence than in its natural environment. In this context, the term "homologous" is intended to include a cDNA sequence encoding a protease native to the host organism in question. The term "heterologous" is intended to include a DNA sequence not expressed by the host cell in nature. Thus, the DNA sequence may be from another organism, or it may be a synthetic sequence.

The host cell of the invention, into which the DNA construct or the recombinant expression vector of the invention is to be introduced, may be any cell which is capable of producing the protease and includes bacteria, yeast, fungi and higher eukaryotic cells.

Examples of bacterial host cells which, on cultivation, are capable of producing the protease are grampositive bacteria such as strains of *Bacillus*, in particular a strain of *Bacillus subtilis, Bacillus licheniformis, Bacillus lentus, Bacillus brevis, Bacillus stearothermophilus, Bacillus alkalophilus, Bacillus amyloliquefaciens, Bacillus coagulans, Bacillus circulans, Bacillus lautus*, *Bacillus megatherium*, *Bacillus pumilus*, *Bacillus thuringiensis* or *Bacillus agaradherens*, or strains of *Streptomyces*, in particular a strain of *Streptomyces lividans* or *Streptomyces murinus*, or gramnegative bacteria such as *Echerichia coli*. The transformation of the bacteria may be effected by protoplast transformation or by using competent cells in a manner known *per se* (cf. e.g. *Sambrook et al.*, Molecular Cloning. A Laboratory Manual, Cold Spring Harbor, NY, 1989).

When expressing the protease in bacteria such as *Escherichia coli*, the protease may be retained in the cytoplasm, typically as insoluble granules (known as inclusion bodies), or may be directed to the periplasmic space by a bacterial secretion

sequence. In the former case, the cells are lysed and the granules are recovered and denatured after which the protease is refolded by diluting the denaturing agent. In the latter case, the protease may be recovered from the periplasmic space by disrupting the cells, e.g. by sonication or osmotic shock, to release the contents of the periplasmic space and recovering the protease.

Examples of suitable yeasts cells include cells of *Saccharomyces* sp., in particular strains of *Saccharomyces cerevisiae*, *Saccharomyces kluyveri*, and *Saccharomyces uvarum*, cells of *Schizosaccharomyces* sp., such as *Schizosaccharomyces pombe*, cells of *Kluyveromyces*, such as *Kluyveromyces lactis*, cells of *Hansenula*, e.g. *Hansenula polymorpha*, cells of *Pichia*, e.g. *Pichia pastoris* (cf. *Gleeson et al.*, J. Gen. Microbiol. 132, 1986, pp. 3459-3465; US 4,882,279), and cells of *Yarrowia sp.* such as *Yarrowia lipolytica*. Methods for transforming yeast cells with heterologous DNA and producing heterologous polypeptides there from are described, e.g. in US 4,599,311, US 4,931,373, US 4,870,008, 5,037,743, and US 4,845,075, all of which are hereby incorporated by reference. Transformed cells are selected by a phenotype determined by a selectable marker, commonly drug resistance or the ability to grow in the absence of a particular nutrient, e.g. leucine. A preferred vector for use in yeast is the POT1 vector disclosed in US 4,931,373. The DNA sequence encoding the protease may be preceded by a signal sequence and optionally a leader sequence, e.g. as described above.

Examples of other fungal cells are cells of filamentous fungi, e.g. *Aspergillus* sp., in particular strains of *Aspergillus japonicus*, *Aspergillus oryzae*, *Aspergillus nidulans* or *Aspergillus niger*, *Neurospora* sp., *Fusarium* sp., in particular strains of *Fusarium oxysporum* or *Fusarium graminearum*, or *Trichoderma* sp.. Fungal cells may be transformed by a process involving protoplast formation and transformation of the protoplasts followed by regeneration of the cell wall in a manner known *per se*. The use of *Aspergillus* sp. for the expression of proteins have been described in e.g., EP 272,277 and EP 230,023. The transformation of *F. oxysporum* may, for instance, be carried out as described by *Malardier et al.*, Gene 1989 **78** 147-156. The use of *Aspergillus* as a host microorganism is described in e.g. EP 238 023, the contents of which are hereby incorporated by reference.

The transformed or transfected host cell described above is then cultured in a suitable nutrient medium under conditions permitting the expression of the protease, after which the resulting protease is recovered from the culture.

The medium used to culture the cells may be any conventional medium suitable for growing the host cells, such as minimal or complex media containing appropriate supplements. Suitable media are available from commercial suppliers or

may be prepared according to published recipes (e.g. in catalogues of the American Type Culture Collection). The protease produced by the cells may then be recovered from the culture medium by conventional procedures including separating the host cells from the medium by centrifugation or filtration, precipitating the proteinaceous components of the supernatant or filtrate by means of a salt, e.g. ammonium sulphate, purification by a variety of chromatographic procedures, e.g. ion exchange chromatography, gelfiltration chromatography, affinity chromatography, or the like, dependent on the type of protease in question.

Method of Producing Proteolytic Enzymes

In a still further aspect, the present invention provides a method of producing an enzyme according to the invention, wherein a suitable host cell, which has been transformed with a DNA sequence encoding the enzyme, is cultured under conditions permitting the production of the enzyme, and the resulting enzyme is recovered from the culture.

The medium used to culture the transformed host cells may be any conventional medium suitable for growing the host cells in question. The expressed protease may conveniently be secreted into the culture medium and may be recovered therefrom by well-known procedures including separating the cells from the medium by centrifugation or filtration, precipitating proteinaceous components of the medium by means of a salt such as ammonium sulphate, followed by chromatographic procedures such as ion exchange chromatography, affinity chromatography, or the like.

Enzyme Preparations

In a still further aspect, the present invention provides an enzyme preparation useful for detergent compositions, said preparation being enriched in a proteolytic enzyme as described above.

The enzyme preparation of the invention may be one which comprises an enzyme of the invention as the major enzymatic component, and may in particular be a mono-component enzyme preparation.

The enzyme preparation may be prepared in accordance with methods known in the art and may be in the form of a liquid or a dry preparation. For instance, the enzyme preparation may be in the form of a granulate or a micro granulate. The enzyme preparation may be stabilized in accordance with methods known in the art.

The enzyme preparation according to the invention may be useful for incorporation into detergent compositions, in the feed and food industry for

hydrolyzing proteinaceous substances, for threatment of leather, and for treatment of wool. The dosage of the enzyme preparation of the invention and other conditions under which the preparation is used may be determined on the basis of methods known in the art.

EXAMPLES

The invention is further illustrated with reference to the following examples which are not intended to be in any way limiting to the scope of the invention as claimed.

MATERIALS AND METHODS

Hybridization Conditions

Suitable hybridization conditions for determining hybridization between an oligonucleotide probe and an "analogous" DNA sequence of the invention may be defined as either low to medium stringency conditions or high stringency conditions. A suitable oligonucleotide probe to be used in the hybridization may be prepared on the basis of the DNA sequence shown in SEQ ID NO: 1, or any sub-sequence thereof.

Low to Medium Stringency

A filter containing the DNA fragments to hybridize is subjected to presoaking in 5x SSC, and prehydbridized for 1 hour at about 40°C in a solution of 20% formamide, 5x Denhardt's solution, 50 mM sodium phosphate, pH 6.8, and 50 μg of denatured sonicated calf thymus DNA. After hybridization in the same solution supplemented with 100 μM ATP for 18 hours at about 40°C, the product is washed three times in 2x SSC at a temperature of about 45°C for 30 minutes.

Molecules to which the oligonucleotide probe hybridizes under these conditions are detected using standard detection procedures (e.g. Southern blotting).

High Stringency Hybridization

A filter containing the DNA fragments to hybridize is subjected to presoaking in 5x SSC, and prehybridized for 1 hour at about 50°C in a solution of 5x SSC, 5x Denhardt's solution, 50 mM sodium phosphate, pH 6.8, and 50 μg of denatured sonicated calf thymus DNA. After hybridization in the same solution supple-

mented with 50 µCi 32-P-dCTP labelled probe for 18 hours at ~50°C, the product is washed three times in 2x SSC, 0.2% SDS at 50°C for 30 minutes.

Molecules to which the oligonucleotide probe hybridizes under these conditions are detected using a x-ray film.

Immunological Cross-Reactivity

Antibodies useful for determining immunological cross-reactivity are prepared using a purified protease obtained from the strain *Nocardiopsis sp.* 10R NRRL 18262. More specifically, antiserum against the protease enzyme are raised by immunizing rabbits (or other rodents) according to the procedure described by *Axelsen N H, et al.* in "A Manual of Quantitative Immunoelectrophoresis", Blackwell Scientific Publications, 1973, Chapter 23, or by *Johnstone A & Thorpe R* in "Immunochemistry in Practice", Blackwell Scientific Publications, 1982 (more specifically p. 27-31).

Purified immunoglobulins may be obtained from the antisera, for example by salt precipitation ($(NH_4)_2\ SO_4$), followed by dialysis and ion exchange chromatography, *e.g.* on DEAE-Sephadex. Immunochemical characterization of proteins may be done either by Ouchterlony double-diffusion analysis [*Ouchterlony O*, in "Handbook of Experimental Immunology", *Weir D M*, Ed., Blackwell Scientific Publications, 1967, pp. 655-706], by crossed immunoelectrophoresis [*Axelsen N H, et al.*, *supra*, Chapters 3 and 4], or by rocket immunoelectrophoresis [*Axelsen N H, et al.*, *supra*, Chapter 2].

Example 1

Cloning and Sequencing the Nocardiopsis 10R Gene

From the strain *Nocardiopsis sp.* 10R NRRL 18262, chromosomal DNA was extracted by standard procedures. The total chromosomal DNA was digested with restriction enzyme BamH1 and size-fractionated fragments 3.5-5.5 kb were cloned into the BamH1 site in pUC19 (cf. e.g. *Sambrook et al.*, Molecular Cloning. A Laboratory Manual, Cold Spring Harbor, NY, 1989).

A number of recombinant colonies were screened by standard hybridization technique (hybridization temperature 60°C; wash temperature 60°C) using the following probe:

5'- GTC/G TGC GCG/C GAG CCG/C GGT/C GAC -3'

A number of positive colonies were identified, including the strain LiH370 containing a plasmid pLiH370 with a 4.5 kb BamH1 fragment containing the 10R gene, as determined by DNA sequencing.

The DNA sequence containing the 10R gene is presented as SEQ ID NO: 1, below. The entire mature protein was deduced to contain 188 amino acids.

SEQUENCE LISTING

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1596 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Nocardiopsis

(B) STRAIN: 10R (NRRL 18262)

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION:900..1463

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
ACGTTTGGTA CGGGTACCGG TGTCCGCATG TGGCCAGAAT GCCCCCTTGC GACAGGGAAC      60

GGATTCGGTC GGTAGCGCAT CGACTCCGAC AACCGCGAGG TGGCCGTTCG CGTCGCCACG     120

TTCTGCGACC GTCATGCGAC CCATCATCGG GTGACCCCAC CGAGCTCTGA ATGGTCCACC     180

GTTCTGACGG TCTTTCCCTC ACCAAAACGT GCACCTATGG TTAGGACGTT GTTTACCGAA     240

TGTCTCGGTG AACGACAGGG GCCGGACGGT ATTCGGCCCC GATCCCCCGT TGATCCCCCC     300

AGGAGAGTAG GGACCCCATG CGACCCTCCC CCGTTGTCTC CGCCATCGGT ACGGGAGCGC     360

TGGCCTTCGG TCTGGCGCTG TCCGGTACCC CGGGTGCCCT CGCGGCCACC GGAGCGCTCC     420

CCCAGTCACC CACCCCGGAG GCCGACGCGG TCTCCATGCA GGAGGCGCTC CAGCGCGACC     480

TCGACCTGAC CTCCGCCGAG GCCGAGGAGC TGCTGGCCGC CCAGGACACC GCCTTCGAGG     540

TCGACGAGGC CGCGGCCGAG GCCGCCGGGG ACGCCTACGG CGGCTCCGTC TTCGACACCG     600
```

```
AGAGCCTGGA ACTGACCGTC CTGGTCACCG ATGCCGCCGC GGTCGAGGCC GTGGAGGCCA      660

CCGGCGCCGG GACCGAGCTG GTCTCCTACG GCATCGACGG TCTCGACGAG ATCGTCCAGG      720

AGCTCAACGC CGCCGACGCC GTTCCCGGTG TGGTCGGCTG GTACCCGGAC GTGGCGGGTG      780

ACACCGTCGT CCTGGAGGTC CTGGAGGGTT CCGGAGCCGA CGTCAGCGGC CTGCTCGCGG      840

ACGCCGGCGT GGACGCCTCG GCCGTCGAGG TGACCACGAG CGACCAGCCC GAGCTCTAC       899

GCC GAC ATC ATC GGT GGT CTG GCC TAC ACC ATG GGC GGC CGC TGT TCG       947
Ala Asp Ile Ile Gly Gly Leu Ala Tyr Thr Met Gly Gly Arg Cys Ser
  1               5                  10                  15

GTC GGC TTC GCG GCC ACC AAC GCC GCC GGT CAG CCC GGG TTC GTC ACC       995
Val Gly Phe Ala Ala Thr Asn Ala Ala Gly Gln Pro Gly Phe Val Thr
             20                  25                  30

GCC GGT CAC TGC GGC CGC GTG GGC ACC CAG GTG ACC ATC GGC AAC GGC      1043
Ala Gly His Cys Gly Arg Val Gly Thr Gln Val Thr Ile Gly Asn Gly
         35                  40                  45

AGG GGC GTC TTC GAG CAG TCC GTC TTC CCC GGC AAC GAC GCG GCC TTC      1091
Arg Gly Val Phe Glu Gln Ser Val Phe Pro Gly Asn Asp Ala Ala Phe
     50                  55                  60

GTC CGC GGT ACG TCC AAC TTC ACG CTG ACC AAC CTG GTC AGC CGC TAC      1139
Val Arg Gly Thr Ser Asn Phe Thr Leu Thr Asn Leu Val Ser Arg Tyr
 65                  70                  75                  80

AAC ACC GGC GGG TAC GCA GCG GTC GCC GGT CAC AAC CAG GCC CCC ATC      1187
Asn Thr Gly Gly Tyr Ala Ala Val Ala Gly His Asn Gln Ala Pro Ile
                 85                  90                  95

GGC TCC TCC GTC TGC CGC TCC GGC TCC ACC ACC GGT TGG CAC TGC GGC      1235
Gly Ser Ser Val Cys Arg Ser Gly Ser Thr Thr Gly Trp His Cys Gly
```

```
        100                 105                 110

ACC ATC CAG GCC CGC GGC CAG TCG GTG AGC TAC CCC GAG GGC ACC GTC      1283
Thr Ile Gln Ala Arg Gly Gln Ser Val Ser Tyr Pro Glu Gly Thr Val
        115                 120                 125

ACC AAC ATG ACC CGG ACC ACC GTG TGC GCC GAG CCC GGC GAC TCC GGC      1331
Thr Asn Met Thr Arg Thr Thr Val Cys Ala Glu Pro Gly Asp Ser Gly
    130                 135                 140

GGC TCC TAC ATC TCC GGC ACC CAG GCC CAG GGC GTG ACC TCC GGC GGC      1379
Gly Ser Tyr Ile Ser Gly Thr Gln Ala Gln Gly Val Thr Ser Gly Gly
145                 150                 155                 160

TCC GGC AAC TGC CGC ACC GGC GGG ACC ACC TTC TAC CAG GAG GTC ACC      1427
Ser Gly Asn Cys Arg Thr Gly Gly Thr Thr Phe Tyr Gln Glu Val Thr
                165                 170                 175

CCC ATG GTG AAC TCC TGG GGC GTC CGT CTC CGG ACC TGATCCCCGC           1473
Pro Met Val Asn Ser Trp Gly Val Arg Leu Arg Thr
        180                 185

GGTTCCAGGC GGACCGACGG TCGTGACCTG AGTACCAGGC GTCCCCGCCG CTTCCAGCGG    1533

CGTCCGCACC GGGGTGGGAC CGGGCGTGGC CACGGCCCCA CCCGTGACCG GACCGCCCGG    1593

CTA                                                                  1596
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 188 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Ala Asp Ile Ile Gly Gly Leu Ala Tyr Thr Met Gly Gly Arg Cys Ser
  1               5                  10                  15

Val Gly Phe Ala Ala Thr Asn Ala Ala Gly Gln Pro Gly Phe Val Thr
             20                  25                  30

Ala Gly His Cys Gly Arg Val Gly Thr Gln Val Thr Ile Gly Asn Gly
         35                  40                  45

Arg Gly Val Phe Glu Gln Ser Val Phe Pro Gly Asn Asp Ala Ala Phe
     50                  55                  60

Val Arg Gly Thr Ser Asn Phe Thr Leu Thr Asn Leu Val Ser Arg Tyr
 65                  70                  75                  80

Asn Thr Gly Gly Tyr Ala Ala Val Ala Gly His Asn Gln Ala Pro Ile
                 85                  90                  95

Gly Ser Ser Val Cys Arg Ser Gly Ser Thr Thr Gly Trp His Cys Gly
            100                 105                 110

Thr Ile Gln Ala Arg Gly Gln Ser Val Ser Tyr Pro Glu Gly Thr Val
        115                 120                 125

Thr Asn Met Thr Arg Thr Thr Val Cys Ala Glu Pro Gly Asp Ser Gly
    130                 135                 140

Gly Ser Tyr Ile Ser Gly Thr Gln Ala Gln Gly Val Thr Ser Gly Gly
145                 150                 155                 160

Ser Gly Asn Cys Arg Thr Gly Gly Thr Thr Phe Tyr Gln Glu Val Thr
                165                 170                 175

Pro Met Val Asn Ser Trp Gly Val Arg Leu Arg Thr
            180                 185
```

CLAIMS

I. A DNA construct comprising a DNA sequence encoding a proteolytic enzyme, which DNA sequence,

(a) comprises the DNA sequence presented as SEQ ID NO: 1; or

(b) comprises a sequence analogue to the DNA sequence presented as SEQ ID NO: 1, which analog DNA sequence either

(i) is homologous to the DNA sequence presented as SEQ ID NO: 1; or

(ii) hybridizes with the same oligonucleotide probe as the DNA sequence presented as SEQ ID NO: 1; or

(iii) encodes a polypeptide which is at least 70% homologous with the polypeptide encoded by the DNA sequence presented as SEQ ID NO: 1; or

(iv) encodes a polypeptide which is immunologically reactive with an antibody raised against a purified protease derived from the strain *Nocardiopsis sp.* 10R NRRL 18262, or encoded by the DNA sequence presented as SEQ ID NO: 1.

II. The DNA construct according to claim 1, in which the DNA sequence encoding the proteolytic enzyme is obtainable from a microorganism.

III. The DNA construct according to claim 2, in which the DNA sequence is obtainable from a filamentous fungus, a yeast or a bacteria.

IV. The DNA construct according to claim 3, in which is the DNA sequence is obtainable from a *Actinomycetes*.

V. The DNA construct according to claim 4, in which is the DNA sequence is obtainable from a strain of *Nocardiopsis*.

VI. The DNA construct according to claim 5, in which is the DNA sequence is obtainable from a strain *Nocardiopsis dassonvillei*, or a strain of *Nocardiopsis sp.* 10R.

VII. The DNA construct according to claim 5, in which is the DNA sequence is obtainable from the strain *Nocardiopsis sp.* 10R NRRL 18262.

VIII. A recombinant expression vector comprising a DNA construct according to any of claims 1-7.

IX. The cell comprising a DNA construct according to any of claims 1-7, or the recombinant expression vector according to claim 8.

X. The cell according to claim 9, which is a bacterial cell.

XI. The cell according to claim 10, which is a strain of *Bacillus*, in particular a strain of *Bacillus subtilis, Bacillus licheniformis, Bacillus lentus, Bacillus brevis, Bacillus stearothermophilus, Bacillus alkalophilus, Bacillus amyloliquefaciens, Bacillus coagulans, Bacillus circulans, Bacillus lautus*, *Bacillus megatherium*, *Bacillus pumilus*, *Bacillus thuringiensis* or *Bacillus agaradherens*, or a strain of *Streptomyces*, in particular a strain of *Streptomyces lividans* or *Streptomyces murinus*, or gramnegative bacteria such as *Echerichia coli*.

XII. The cell according to claim 9, which is a eukaryotic cell, in particular a fungal cell, such as a yeast cell or a filamentous fungal cell.

XIII. The cell according to claim 12, which is a strain of *Aspergillus*, in particular a strain of *Aspergillus japonicus*, a strain of *Aspergillus oryzae*, a strain of *Aspergillus nidulans*, or a strain of *Aspergillus niger*, or a strain of *Neurospora* sp., or a strain of *Fusarium* sp., in particular strains of *Fusarium oxysporum* or *Fusarium graminearum*, or a strain of *Trichoderma* sp..

XIV. A method of producing a proteolytic enzyme, the method comprising culturing a cell according to any of claims 9-13 under conditions permitting the production of the enzyme, and recovering the enzyme from the culture.

XV. A proteolytic enzyme, which

(a) is encoded by a DNA construct according to any of claims 1-7;

(b) produced by the method according to claim 14; and/or

(c) is immunologically reactive with an antibody raised against a purified protease derived from the strain *Nocardiopsis sp.* 10R NRRL 18262, or encoded by the DNA sequence presented as SEQ ID NO: 1.

TITLE: NOVEL DNA SEQUENCES

ABSTRACT

This invention relates to novel DNA constructs encoding proteolytic enzymes, as well as recombinant expressions vectors and host cells comprising these DNA constructs, and methods of producing a proteolytic enzyme.